



# **“WQX Best Practices”**

## **Dos and Don'ts for Water Quality Data Destined for WQX / WQX Web**

STORET User Call October 23, 2014

Alex Heppner Gold Systems, Inc.





# Agenda

- We will hold the monthly STORET User call on Thursday October 23, 2014 at 12 noon Eastern using:
  - Call number 866-299-3188, access 202-566-0399#
  - Please remember send an email to [storet@epa.gov](mailto:storet@epa.gov) to indicate that you are on the call.
  - Webinar link: <https://epa.connectsolutions.com/storetusercall/>
- Agenda
  - Common data reporting errors – As many monitoring activities wind down and attention turns to data submission, we will present on common WQX reporting errors.
  - Other topics and questions



# Attendees

- There were 24 folks on the webinar, the following submitted their names to the storet email indicating that they were on the call:
  - Diane Stevenson - Cherokee Nation Environmental Programs
  - James Christman
  - Kris Dozark - South Dakota Department of Environment and Natural Resources
  - Melanie Titus - NHDES, Watershed Management Bureau
  - Michael Whitman - West Virginia Department of Environmental Protection



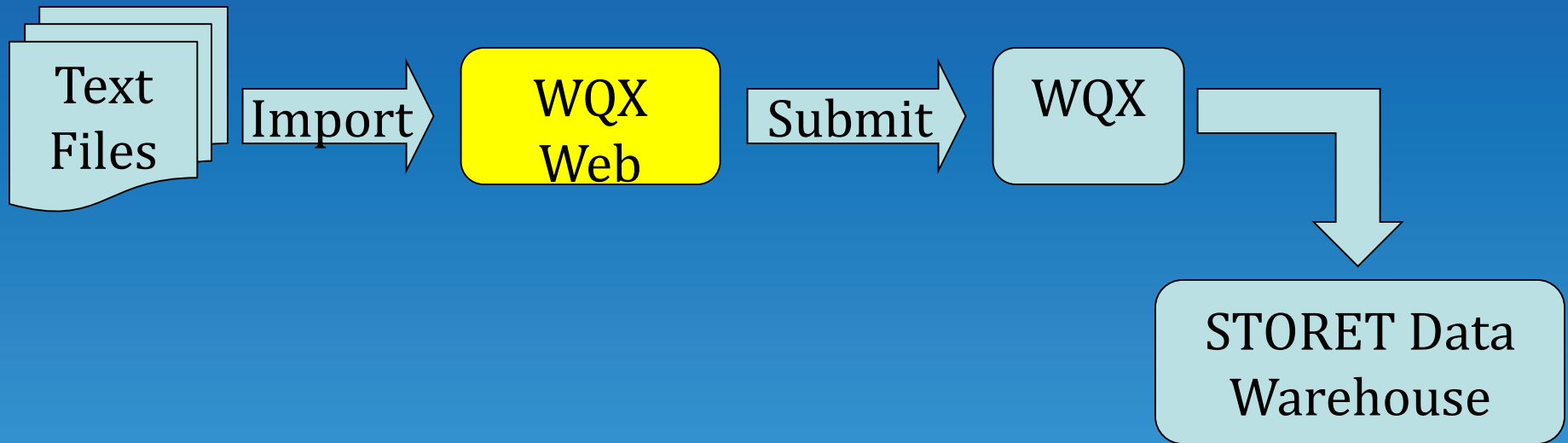
# CWA 106 Grant Funding

- “As a grant requirement for continuing to receive Section 106 funds, tribes must submit to their EPA regional offices annual tribal Assessment Reports that consist of...electronic copies of surface water quality data for nine basic parameters”  
(Final CWA Section 106 Tribal Grant Guidance)
- “WQX is the new way to place data into the STORET Warehouse.”  
(Final CWA Section 106 Tribal Grant Guidance - Data Management)



# Typical WQX Flow

## Using the EPA “WQX Web” Submission Tool





# Based on Experience...

Created and support:

- Water Quality eXchange (WQX) for the EPA
- “WQX Web” data submission tool for the EPA



# Common Issues within Datasets

Let's take a look at common issues and their solutions when preparing to submit via WQX Web:



# Unorganized Data - Solutions

## **Get a database!**

Spreadsheets are not a database

Stacks of paper are not a database

All data should be stored in a single location and format:

- Can be hosted in a “Cloud” environment, or on a local server.
- Accessible by many (or few, depending on Organizational preference)
  - Secure with various security levels/permissions
- Only one system to backup
- Easy to search on and access data





# 1. Data not provided in an electronic format

- Labs sometimes provide results on paper or electronically in Adobe Portable Document Format (PDF).
- Labs can save their customers significant work by providing result data as either a comma (.csv), tab delimited text file (.txt), or as a spreadsheet (e.g. Microsoft Excel .xlsx, .xls).
- If your lab can't or won't, you might want to shop around.



# Be-Aware of Excel

Excel will convert anything that looks like a number into a number.

- To the extent possible, avoiding using numeric IDs that have preceding zeros (that are significant).
  - For example, a Monitoring Location ID of "032" may get changed to "32" by Excel.
  - An ID of "32" or "ML032" would avoid data corruption by Excel
- HUICS that start with zero will lose the zero in Excel
  - "03150202" will become "3150202"
  - (The next release of WQX Web will automatically fix HUICS like this.)

Leading zeros and decimal places should always be double checked within a dataset before submission to WQX



## 2. White space and/or “divider” rows or columns

- Data should be in delimited rows and columns with no white space or empty rows.
- “Divider” rows or columns within the result data.
- Row may help a human reader to understand or find particular data, but the row is not result data.
- The lab EDD (Electronic Data Deliverable) should not include any rows or columns in the data section of the file that are not strictly result data and associated meta-data.



## 2. White space and/or “divider” rows or columns


Analyte	Results	Units	R.L	Method	Date	Analyst
Calcium	<u>208</u>	mg/L	0.5 mg/L	SM3111-b	8/1/2011	CS
Magnesium	<u>50.5</u>	mg/L	0.5 mg/L	SM3111-b	8/1/2011	CS
Iron	<u>0.8</u>	mg/L	0.1 mg/L	SM3111-b	8/1/2011	CS
Potassium	<u>20.7</u>	mg/L	0.5 mg/L	SM3111-b	8/1/2011	CS
Manganese	<u>0.2</u>	mg/L	0.05 mg/L	SM3111-b	8/1/2011	CS
Sodium	<u>281</u>	mg/L	0.5 mg/L	SM3111-b	8/1/2011	AR
Sodium Absorption Ratio	<u>4.5</u>	-	N/A	USDA 20b	8/1/2011	Calculated
Total Hardness as CaCO <sub>3</sub>	<u>727</u>	mg/L	N/A	SM2340-B	8/1/2011	Calculated
Hardness in grains/gallon	<u>42.5</u>	gr/gal	N/A	SM2340-B	8/1/2011	Calculated
Orthophosphate as P	<u>0.1</u>	mg/L	0.1 mg/L	SM4500-P	7/22/2011	AR
Fluoride	<u>0.4</u>	mg/L	0.1 mg/L	SM4500-F	7/24/2011	AR
Sulfate	<u>1100</u>	mg/L	5.0 mg/L	SM4500-SO <sub>4</sub> <sup>2-</sup> E	8/15/2011	CS
Chloride	<u>15.1</u>	mg/L	1.0 mg/L	SM4500-Cl <sup>-</sup>	7/23/2011	AR
Nitrate+Nitrite as N	<u>&lt;0.1</u>	mg/L	0.1 mg/L	EPA 353.2	8/1/2011	AR
Ammonia Nitrogen	<u>&lt;0.1</u>	mg/L	0.1 mg/L	EPA 350.1	6/21/2011	AR
Anion Summation	<u>29.8</u>	meq/L	N/A	SM1030-F	8/1/2011	Calculated
Cation Summation	<u>27.3</u>	meq/L	N/A	SM1030-F	8/1/2011	Calculated
Total Dissolved Solids; Summation	<u>1871</u>	mg/L	N/A	SM1030-F	8/1/2011	Calculated
Ion Summation; Percent Error	<u>4.35</u>	%	N/A	SM1030-F	8/1/2011	Calculated
Total Suspended Solids	<u>240</u>	mg/L	1 mg/L	SM2540-D	8/1/2011	CS
Conductivity	<u>2330</u>	µS/cm	N/A	SM2510-b	7/21/2011	MLB
pH	<u>8.37</u>	-	N/A	SM4500-H <sup>+</sup> b	8/1/2011	AR
Phenolphthalein Alkalinity	<u>6</u>	mg/L CaCO <sub>3</sub>	4.0 mg/L	SM2320-b	8/1/2011	AR
Carbonate	<u>12</u>	mg/L CaCO <sub>3</sub>	4.0 mg/L	SM2320-b	8/1/2011	AR
Bicarbonate	<u>312</u>	mg/L CaCO <sub>3</sub>	4.0 mg/L	SM2320-b	8/1/2011	AR
Hydroxide	<u>0</u>	mg/L CaCO <sub>3</sub>	0 mg/L	SM2320-b	8/1/2011	AR
Carbonate	<u>7.2</u>	mg/L CO <sub>3</sub>	4.0 mg/L	SM2320-b	8/1/2011	AR
Bicarbonate	<u>381</u>	mg/L HCO <sub>3</sub>	4.0 mg/L	SM2320-b	8/1/2011	AR
Hydroxide	<u>0</u>	mg/L OH	0 mg/L	SM2320-b	8/1/2011	AR
Total Alkalinity	<u>324</u>	mg/L CaCO <sub>3</sub>	4.0 mg/L	SM2320-b	8/1/2011	AR
Trace Metal Analysis						
	Results	Units	R.L.	Method	Date	Analyst
Cadmium	<u>&lt;0.25</u>	µg/L	0.25 ug/L	SM3113b	8/3/2011	AR
Selenium	<u>&lt;2.0</u>	µg/L	2.0 ug/L	SM3113b	8/2/2011	AR
Arsenic	<u>&lt;2.0</u>	µg/L	2.0 ug/L	SM3113b	9/1/2011	AR
Lead	<u>&lt;2.0</u>	µg/L	2.0 ug/L	SM3113b	8/10/2011	AR
Microbiological Analysis						
	Results	Units	Method	Date	Analyst	
Total Coliform	<u>&gt;2419</u>	MPN/100ml	N/A	SM9223b	7/22/2011	AR



### 3. Date and Time values included in a single column

- WQX reads Date and Time as two distinct data elements.
- Any date and time data should be listed in separate columns, as shown below. NOTE: the time does not have to be in a 24hr format.

8/29/2013 3:40:00 PM			
	D	E	
D	DEPTNAME	COLLECTDATE	P
		8/29/2013	
		8/29/2013	
		9/4/2013	
		9/4/2013	
		8/29/2013	
		8/29/2013	

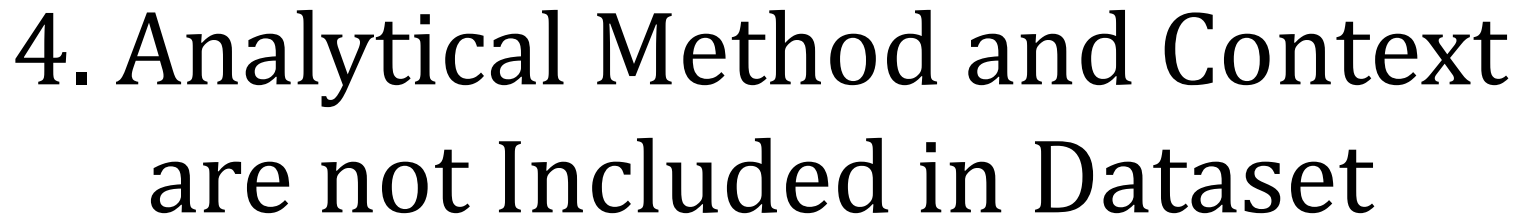
A large red prohibition sign (a circle with a diagonal slash) is overlaid on the table, indicating that combining date and time in a single column is incorrect. A dashed arrow points from the sign to the 'COLLECTDATE' header and the date values in the table.

E	F
COLLECTDATE	COLLECTTIME
8/29/2013	15:40:00
8/29/2013	15:40:00
9/4/2013	15:20:00
9/4/2013	15:20:00
8/29/2013	15:01:00
8/29/2013	15:01:00
9/4/2013	13:10:00
9/4/2013	13:10:00
8/29/2013	15:01:00
8/29/2013	15:01:00
8/28/2013	15:37:00
8/28/2013	15:37:00
8/28/2013	10:13:00



## 4. Analytical Method and Context are not Included in Dataset

- WQX requires the Analytical Method ID and the Method Context for results from a sampling activity.
- The Analytical Method and the associated Context should be in separate columns.

[illegible]



## 5. Detection Conditions and Limits with Result Value.

- WQX tracks result values separately from Detection Conditions.
- WQX defines “Detection Condition” as scenarios when an actual result value cannot be provided.
- Result Value is the actual measurement for each characteristic.





## 5. Detection Conditions and Limits with Result Value.

- Either the Result Value field or the Result Detection Condition field must contain a value; never both fields.
- **Example:** If the result is below detection limit, the Result Value must be blank, and the Result Detection Condition field must contain 'Not Detected'.



## 5. Detection Conditions and Limits with Result Value.

If you report a value of “Not Detected”:

- Provide the Detection Limit in the Result Detection/Quantitation Limit Measure field. This should be provided by the lab along with appropriate units.
- Provide the Detection Limit Units in the Result Detection Quantitation Limit Unit field. The units should be the same as what the Result Value Units would be.
- Provide the Result Detection/Quantitation Limit Type, which is a choice list for why the analyte was not detected or quantified.



# 5. Detection Conditions and Limits with Result Value.

Characteristic Name	Method Speciation Name	Result Sample Fraction	Result Measure	Result Unit	Result Detection Condition	Result Detection Quantitation Limit Measure	Result Detection Quantitation Limit Unit	Result Detection/Quantitation Limit Type	Result Stat
Nitrate	as N	T	0.49	mg/l					
Phosphorus		T	0.127	mg/l					
Ammonia-nitrogen		T	0.3	mg/l					
Kjeldahl nitrogen		T	1.6	mg/l					
Hardness, carbonate	as CaCO <sub>3</sub>	T	425	mg/l					
Calcium		T	94.3	mg/l					
Copper		T			Not Detected		0.005 mg/l	Lower Reporting Limit	
Iron		T	0.853	mg/l					
Lead		T			Not Detected		0.03 mg/l	Lower Reporting Limit	
Magnesium		T	46.1	mg/l					
Manganese		T	0.512	mg/l					
Sodium		T	20.7	mg/l					
Nitrite	as N	T	0.056	mg/l					
Total dissolved solids		T	606	mg/l					



## 6. What's an Activity?

Required for any sort of data in WQX, an Activity ID can be described as such:

A sampling event taken at:

- One site
- On a specific date
- At a specific time
- With a specific Activity Type

Multiple samples can be associated with the sample Activity, as long as they are within the same Activity Type.

How do I create an effective Activity ID?



## 6. What's an Activity?

Create a function in Excel!

Make Excel do the work by creating a function that draws from other data elements within your dataset

For Example:

Site:Date:Time:ActivityType

Example of an excel function:

```
=CONCATENATE(M3,":",TEXT(B3,"YYYYMMDD"),":",TEXT(D3,"HHMM"),":F")
```



## 7. Activity Type Not Clearly Identified

- WQX require that QC sample results have a type specifying what category of QC sample was taken and analyzed.
- Include a specific column for Activity Type in the results file.

-For a list of all allowed activity types see:

[http://www.epa.gov/storet/wqx/tutorial/Use\\_lookup\\_tables.html](http://www.epa.gov/storet/wqx/tutorial/Use_lookup_tables.html)



## 7. Activity Type Not Clearly Identified

	A	B	C
1	Activity ID	Activity Type	Activity Media Name
4877	072011-11	Sample-Routine	Water
4878	072011-11	Sample-Routine	Water
4879	072011-11	Sample-Routine	Water
4880	072011-11	Sample-Routine	Water
4881	072011-11	Sample-Routine	Water
4882	072011-11D	Quality Control Sample-Lab Duplicate	Water
4883	072011-11D	Quality Control Sample-Lab Duplicate	Water
4884	072011-11D	Quality Control Sample-Lab Duplicate	Water
4885	072011-11D	Quality Control Sample-Lab Duplicate	Water
4886	072011-11D	Quality Control Sample-Lab Duplicate	Water
4887	072011-11D	Quality Control Sample-Lab Duplicate	Water
4888	072011-11D	Quality Control Sample-Lab Duplicate	Water



## 8. Sample Fraction and/or Method Speciation not clearly identified

- The EDD (Electronic Data Deliverable) should list Characteristic Name, the Sample Fraction, and the Method Speciation in separate columns.

	K	L	M	N	O
1	CHARACTERISTIC NAME	METHOD SPECIATION	SAMPLE FRACTION	MATRIX	METHOD
20	Mercury		Dissolved	SW	M245.1 CVAA
21	Potassium		Dissolved	SW	M200.7 ICP
22	Selenium		Dissolved	SW	M200.8 ICP-MS
23	Selenium		Total	SW	M200.8 ICP-MS
24	Silver		Total	SW	M200.7 ICP
25	Silver		Dissolved	SW	M200.7 ICP
26	Sodium		Dissolved	SW	M200.7 ICP
27	Bicarbonate as CaCO3	as CaCO3		SW	SM2320B - Titration
28	Carbonate as CaCO3	as CaCO3		SW	SM2320B - Titration
29	Cation-Anion Balance			SW	Calculation
30	Chemical Oxygen Demand			SW	M410.4
31	Chloride			SW	SM4500Cl-E
32	Conductivity @25C			SW	SM2510B
33	Hardness as CaCO3	as CaCO3		SW	SM2340B - Calculation
34	Hydroxide as CaCO3	as CaCO3		SW	SM2320B - Titration
35	Nitrate as N	as N		SW	Calculation: NO3
36	Nitrate/Nitrite as N	as N		SW	M353.2 - H2SO4 p
37	Nitrite as N	as N		SW	M353.2 - Automar
38	Nitrogen ammonia			SW	M350.1 - Automar
39	Nitrogen Kjeldahl			SW	M351.2 - TKN by E





## 9. Values don't match the WQX spelling

- The EDD should follow the standardized WQX lists of allowed values.
- A spreadsheet of the WQX allowed value tables is available for download:

<http://cdx.epa.gov/wqx/download/DomainValues/All.zip>



# 10. Data Entry Errors

- With manual data entry (into your dataset), there is a higher propensity for error due to typos and errors in transcription.
- When possible, mitigate this by importing the electronic files generated at the source (e.g. from your sonde, from your lab, etc.) into your database.
- Sometimes, even the source files have data entry errors (Lab enters an incorrect result value in a spreadsheet). For Example: wrong UoM set up in a Field Measurement/Probe configuration.



# 10. Data Entry Errors

Your data management solution needs to catch errors before they get into the system:

- Set up QA/QC thresholds per Characteristic, Method Speciation, Unit of Measure, and Sample Fraction Combination.
- Validate result values against your thresholds during both batch import and manual data entry.
- Configure default values for numerous data elements associated with an Analyte result value. (Saves time / reduces errors)

These are all checks a Laboratory and/or Database can do for your Organization!



# 11. Monitoring Location Management

- Maintain a consistent list of monitoring locations in your local, or “cloud” database.
- Limit who can create new monitoring locations to avoid duplicates.
- Use GIS tools to identify duplicate locations.
- Develop Procedures for merging duplicate Monitoring Locations.



# 11. Monitoring Location Management

## Duplicate Locations:

- Programs often sample at the same physical Location under multiple names.
- Over time, new samplers return to old sites but give them new names.
- Multiple programs (106 vs 319) may sample at same location under different names.
- Often are same Location with slightly different Lat/Long recorded.



# Review

- Data Organization is key to avoid issues further down the road, whether it be through a “Cloud” database service, or locally hosted database.
- Avoid combining fields that will create invalid column formats for importing into WQX Web.
  - While it may “look” better, or be easily readable, it’s important to stick to creating a “WQX” compatible format.
- Once you decide on an appropriate spreadsheet format, consult with your laboratory to request they report your data in the preferred format



# Questions? Please Contact Us

Charles Kovatch  
(202) 566-0399  
[kovatch.charles@epa.gov](mailto:kovatch.charles@epa.gov)

Laura Shumway  
(202) 566-2514  
[Shumway.Laura@epa.gov](mailto:Shumway.Laura@epa.gov)

Mark LeBaron  
(801) 456-6145  
[markl@goldsystems.com](mailto:markl@goldsystems.com)  
[www.awqms.com](http://www.awqms.com)

Alex Heppner  
(801) 456-6133  
[alexh@goldsystems.com](mailto:alexh@goldsystems.com)  
[www.awqms.com](http://www.awqms.com)